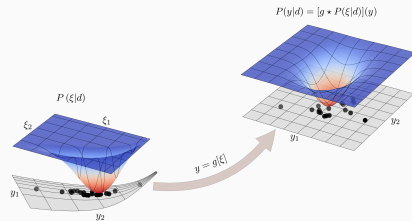


Geometric Variational Inference

USING RIEMANNIAN GEOMETRY TO ENHANCE VARIATIONAL METHODS

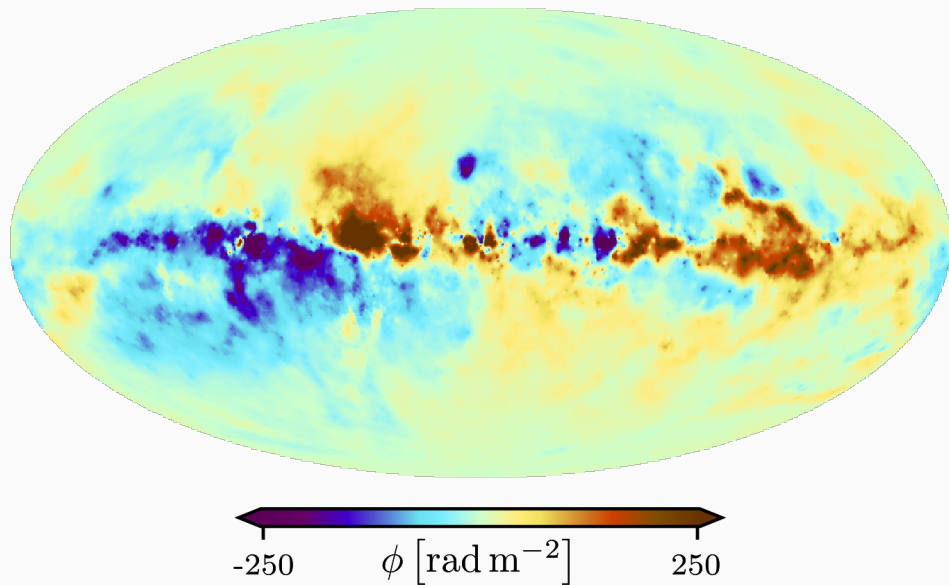


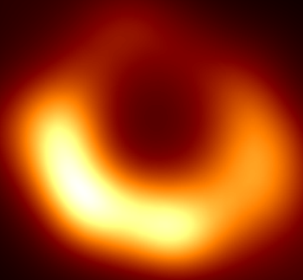
Philipp Frank¹, Reimar Leike³, Torsten Ensslin^{1,2}

Seminar: Signal reconstruction with python, MPA, Garching, September 21, 2022

- (1) Max-Planck Institute for Astrophysics MPA, Garching, Germany
- (2) Ludwig-Maximilians University LMU, Munich, Germany
- (3) Amazon Development Center, Cambridge, United Kingdom









Product Rule of Probabilities aka Bayes' theorem

$$\mathcal{P}(s|d, \mathcal{M}) = \frac{\mathcal{P}(d|s, \mathcal{M}) \mathcal{P}(s|\mathcal{M})}{\mathcal{P}(d|\mathcal{M})}$$

Definitions: s := parameters, d := data, $\mathcal{M}$: model assumptions.

Product Rule of Probabilities aka Bayes' theorem

$$\mathcal{P}(s|d, \mathcal{M}) = \frac{\mathcal{P}(d|s, \mathcal{M}) \mathcal{P}(s|\mathcal{M})}{\mathcal{P}(d|\mathcal{M})}$$

Definitions: s := parameters, d := data, $\mathcal{M}$: model assumptions.

Generative prior model: $s(\xi) = F_M(\xi)$ with $P(\xi) = \mathcal{N}(\xi|0, \mathbb{1})$

Product Rule of Probabilities aka Bayes' theorem

$$\mathcal{P}(\xi|d, \mathcal{M}) = \frac{\mathcal{P}(d|s(\xi), \mathcal{M})\mathcal{N}(\xi|0, \mathbf{1})}{\mathcal{P}(d|\mathcal{M})}$$

Definitions: ξ := parameters, d := data, $\mathcal{M}$: model assumptions.

Generative prior model: $s(\xi) = F_M(\xi)$ with $P(\xi) = \mathcal{N}(\xi|0, \mathbf{1})$

Variational Inference

Kullback-Leibler divergence

$$\text{KL}[\mathcal{Q}_\sigma || \mathcal{P}] = - \int \log \left(\frac{\mathcal{P}(\xi|d)}{\mathcal{Q}_\sigma(\xi)} \right) \mathcal{Q}_\sigma(\xi) \, d\xi$$

Posterior: $\mathcal{P}(\xi|d)$; Approximation: $\mathcal{Q}_\sigma(\xi)$; Variational parameters: σ .

Variational Inference (VI)

Kullback-Leibler divergence

$$\text{KL}[\mathcal{Q}_\sigma || \mathcal{P}] = - \int \log \left(\frac{\mathcal{P}(\xi|d)}{\mathcal{Q}_\sigma(\xi)} \right) \mathcal{Q}_\sigma(\xi) \, d\xi$$

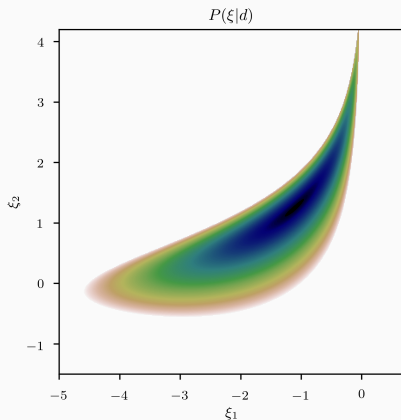
Posterior: $\mathcal{P}(\xi|d)$; Approximation: $\mathcal{Q}_\sigma(\xi)$; Variational parameters: σ .

Approximate distribution $\mathcal{Q}$: $\mathcal{Q}(\mathbf{y}) = \mathcal{N}(\mathbf{y}; 0, \mathbf{1})$

Coordinate system $\mathbf{y} = g_\sigma(\xi)$ such that the *posterior* is close to Normal.

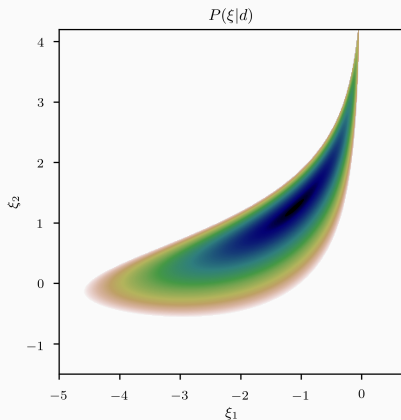
Geometric Variational Inference

Geometric Variational Inference (geoVI) [FLE21]



Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log(\mathcal{P}(\xi|d))$

Geometric Variational Inference (geoVI) [FLE21]

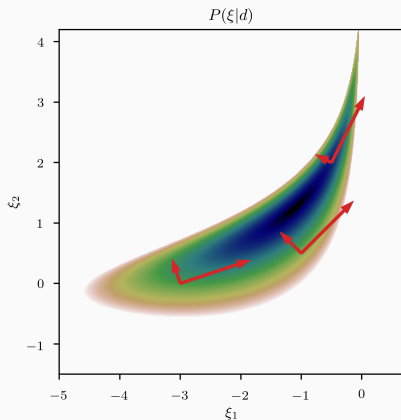


Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log(\mathcal{P}(\xi|d))$

Posterior metric $\mathcal{M}(\xi)$: $\mathcal{M}_{\text{lh}}(\xi) + \mathbb{1}$

Fisher information metric $\mathcal{M}_{\text{lh}}(\xi)$: $\left\langle \frac{\partial^2 \mathcal{H}(d|\xi)}{\partial \xi \partial \xi'} \right\rangle_{\mathcal{P}(d|\xi)}$

Geometric Variational Inference (geoVI) [FLE21]

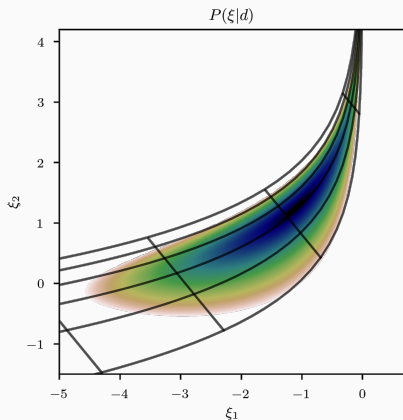


Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log(\mathcal{P}(\xi|d))$

Posterior metric $\mathcal{M}(\xi)$: $\mathcal{M}_{\text{lh}}(\xi) + \mathbb{1}$

Fisher information metric $\mathcal{M}_{\text{lh}}(\xi)$: $\left\langle \frac{\partial^2 \mathcal{H}(d|\xi)}{\partial \xi \partial \xi'} \right\rangle_{\mathcal{P}(d|\xi)}$

Geometric Variational Inference (geoVI) [FLE21]

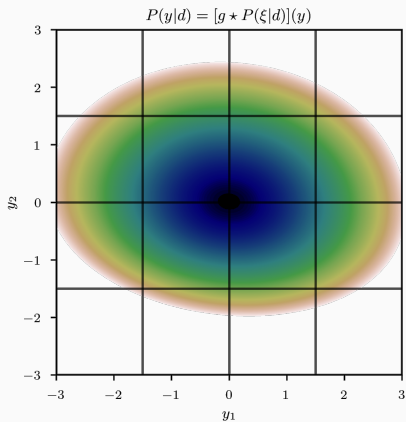
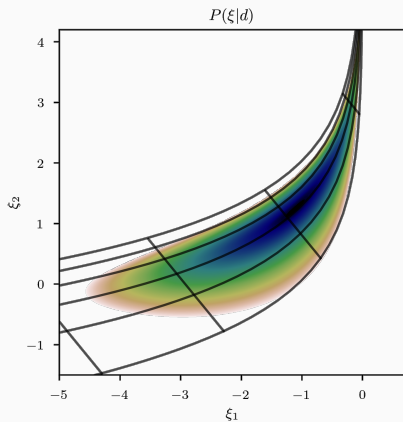


Information Hamiltonian $\mathcal{H}(\xi|d)$: $-\log(\mathcal{P}(\xi|d))$

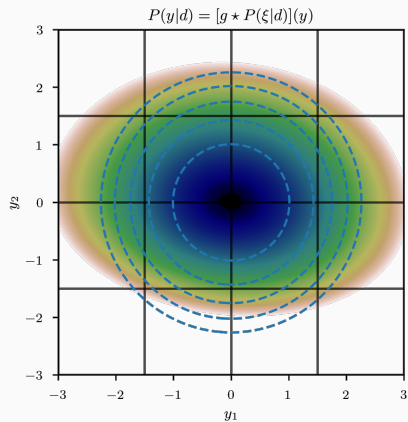
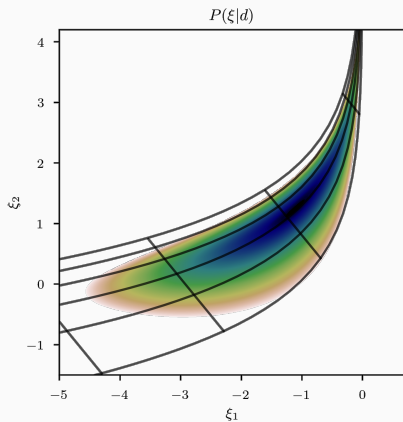
Posterior metric $\mathcal{M}(\xi)$: $\mathcal{M}_{\text{lh}}(\xi) + \mathbb{1}$

Fisher information metric $\mathcal{M}_{\text{lh}}(\xi)$: $\left\langle \frac{\partial^2 \mathcal{H}(d|\xi)}{\partial \xi \partial \xi'} \right\rangle_{\mathcal{P}(d|\xi)}$

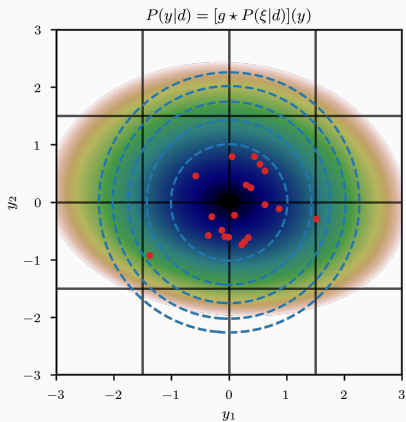
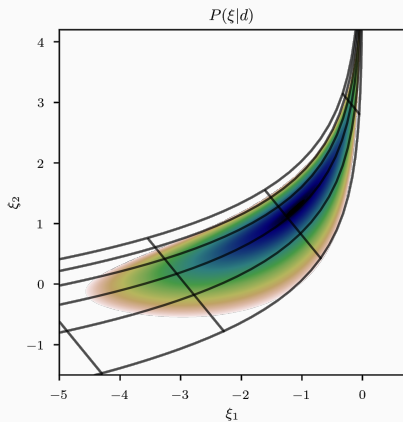
Geometric Variational Inference (geoVI) [FLE21]



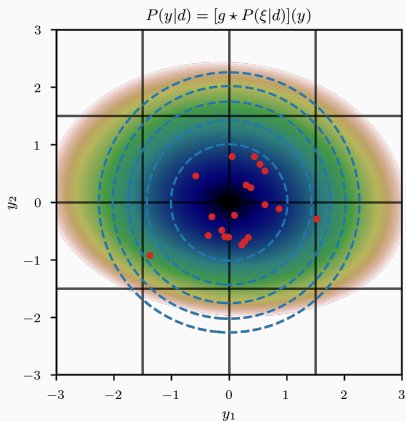
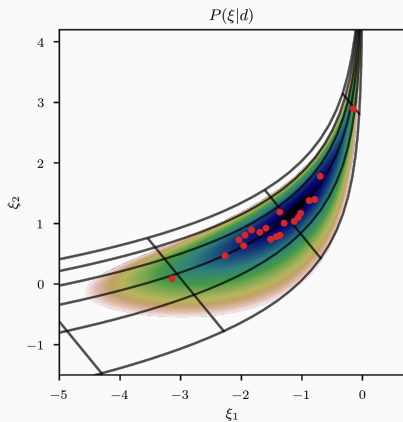
Geometric Variational Inference (geoVI) [FLE21]



Geometric Variational Inference (geoVI) [FLE21]



Geometric Variational Inference (geoVI) [FLE21]



- † Approximate coordinate transformation g such that $\mathcal{M}(\xi) \approx \left(\frac{\partial g}{\partial \xi}\right)^T \frac{\partial g}{\partial \xi}$
- † Fisher metric via pullback: $\mathcal{M}(\xi) = \mathcal{M}_{\text{lh}}(\xi) + \mathbb{1} = \left(\frac{\partial x}{\partial \xi}\right)^T \frac{\partial x}{\partial \xi} + \mathbb{1}.$

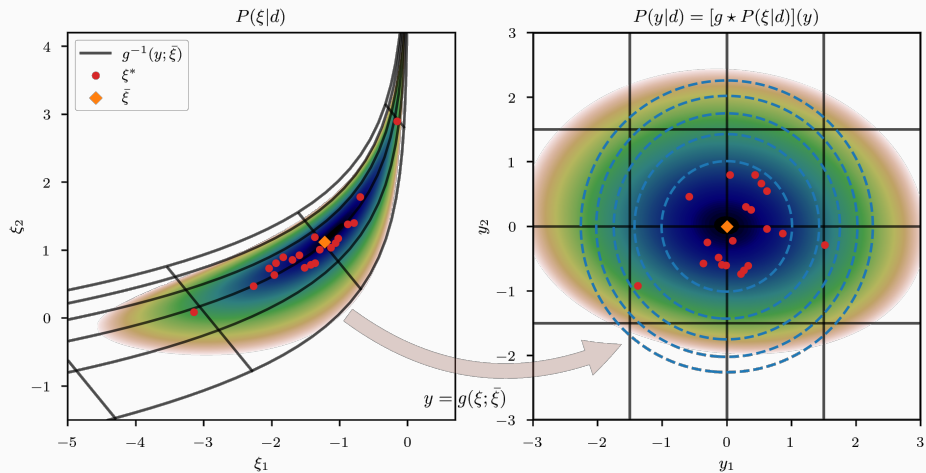
- † Approximate coordinate transformation g such that $\mathcal{M}(\xi) \approx \left(\frac{\partial g}{\partial \xi}\right)^T \frac{\partial g}{\partial \xi}$
- † Fisher metric via pullback: $\mathcal{M}(\xi) = \mathcal{M}_{\text{lh}}(\xi) + \mathbb{1} = \left(\frac{\partial x}{\partial \xi}\right)^T \frac{\partial x}{\partial \xi} + \mathbb{1}$.

Local Euclidean isometry around $\bar{\xi}$

$$y = g(\xi; \bar{\xi}) = \mathcal{M}(\bar{\xi})^{-1/2} \left[\xi - \bar{\xi} + \left(\frac{\partial x}{\partial \xi} \right)^T \Big|_{\xi=\bar{\xi}} (x(\xi) - x(\bar{\xi})) \right]$$

Likelihood transformation: $x(\xi) = x(s(\xi))$, expansion point: $\bar{\xi}$.

Geometric Variational Inference (geoVI) [FLE21]



Local Euclidean isometry around $\bar{\xi}$

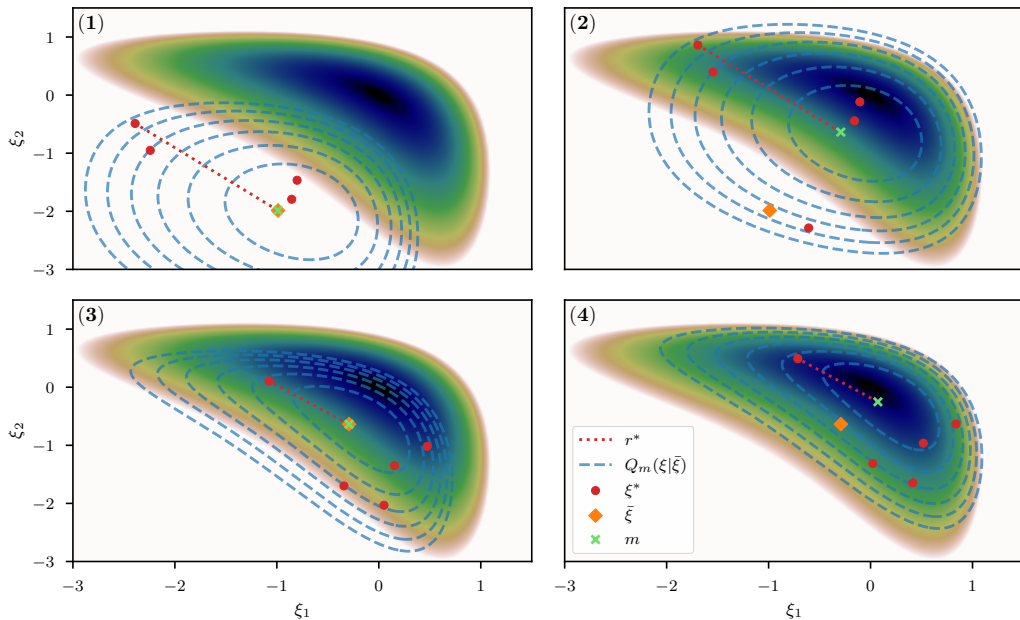
$$y = g(\xi; \bar{\xi}) = \mathcal{M}(\bar{\xi})^{-1/2} \left[\xi - \bar{\xi} + \left(\frac{\partial x}{\partial \xi} \right)^T \bigg|_{\xi=\bar{\xi}} (x(\xi) - x(\bar{\xi})) \right]$$

Likelihood transformation: $x(\xi) = x(s(\xi))$, expansion point: $\bar{\xi}$.

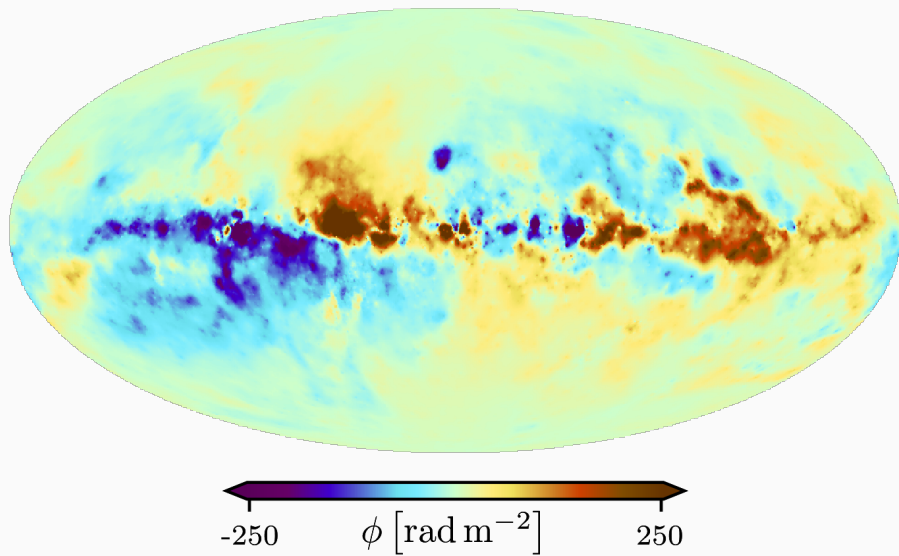
Variational approximation with transformed distribution $\mathcal{Q}$

$$\mathcal{Q}_{\bar{\xi}}(\xi) = \mathcal{N}(y|0, \mathbb{1})|_{y=g(\xi; \bar{\xi})} \left\| \frac{\partial g(\xi; \bar{\xi})}{\partial \xi} \right\|$$

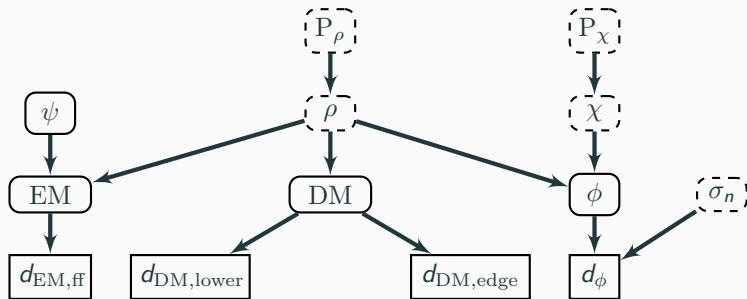
Geometric Variational Inference (geoVI) [FLE21]



Application



Application





Numerical information field theory

<https://gitlab.mpcdf.mpg.de/ift/nifty>

 Philipp Arras, Philipp Frank, Philipp Haim, Jakob Knollmüller, Reimar Leike, Martin Reinecke, and Torsten Enßlin.

Variable structures in m87* from space, time and frequency resolved interferometry.

Nature Astronomy, 6(2):259–269, 2022.

 Philipp Frank, Reimar Leike, and Torsten A. Enßlin.

Geometric variational inference.

Entropy, 23(7), 2021.

 Sebastian Hutschenreuter and Torsten A. Enßlin.

The galactic faraday depth sky revisited.

A&A, 633:A150, 2020.

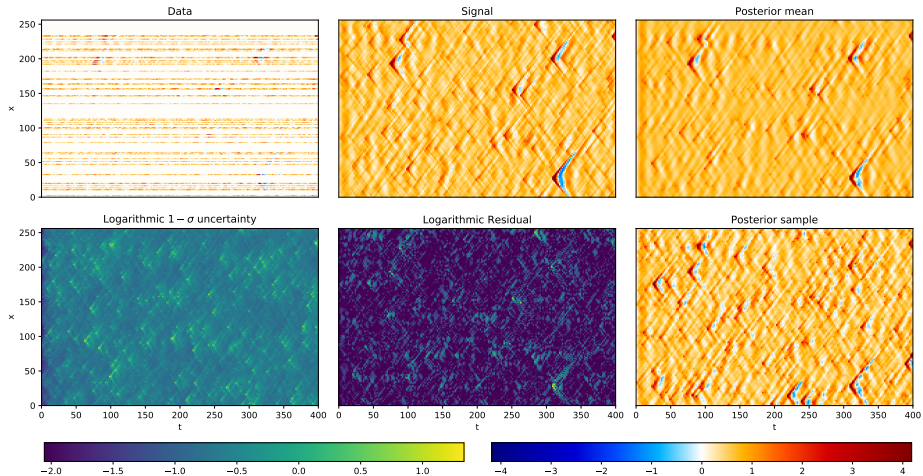
 Reimar Leike, Gordian Edenhofer, Jakob Knollmüller, Christian Alig, Philipp Frank, and Torsten A. Enßlin.

The galactic 3d large-scale dust distribution via gaussian process regression on spherical coordinates.

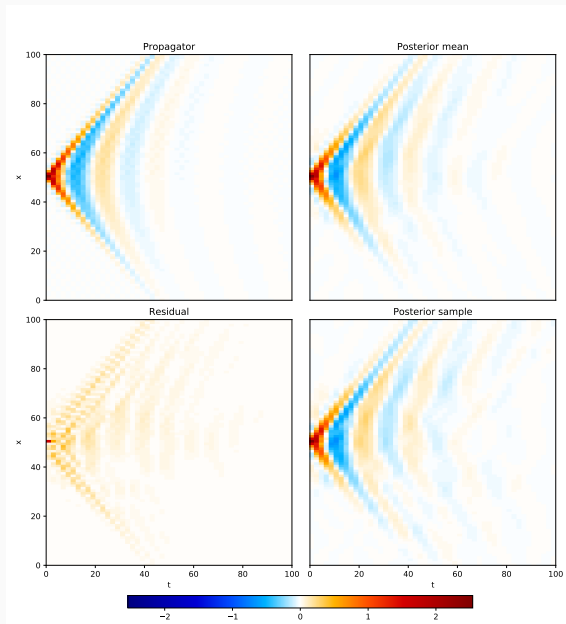
arXiv, 2204.11715, 2022.

Appendix

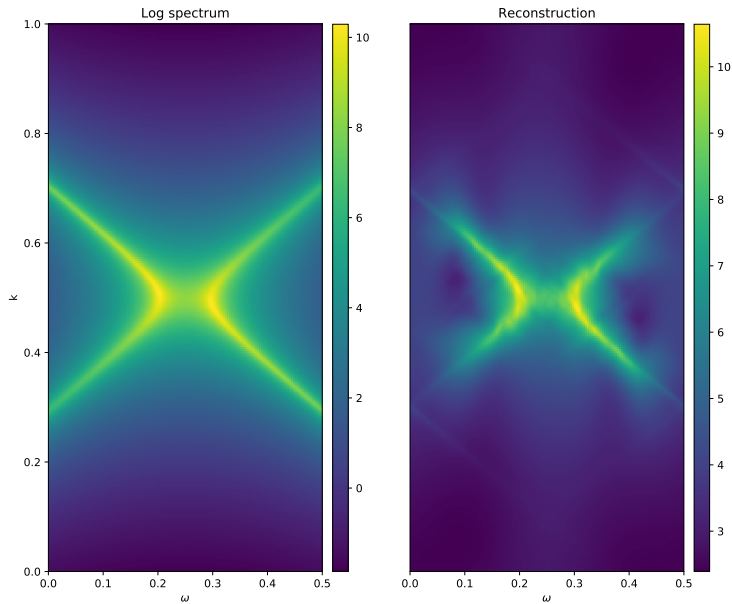
Application



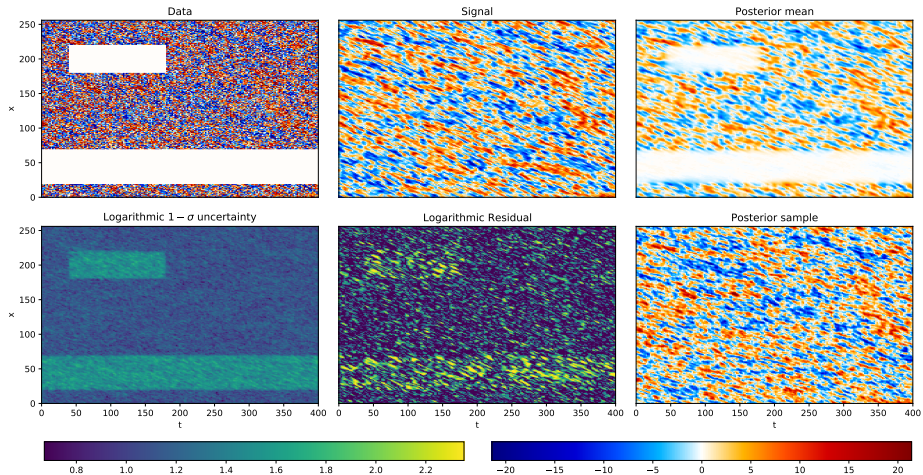
Application



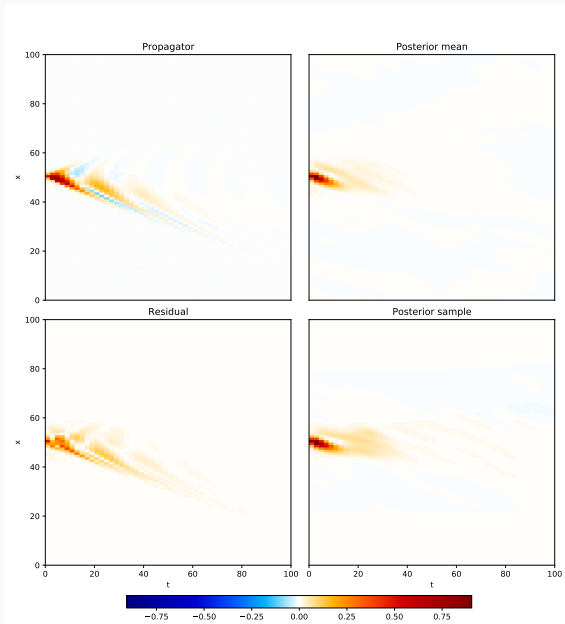
Application



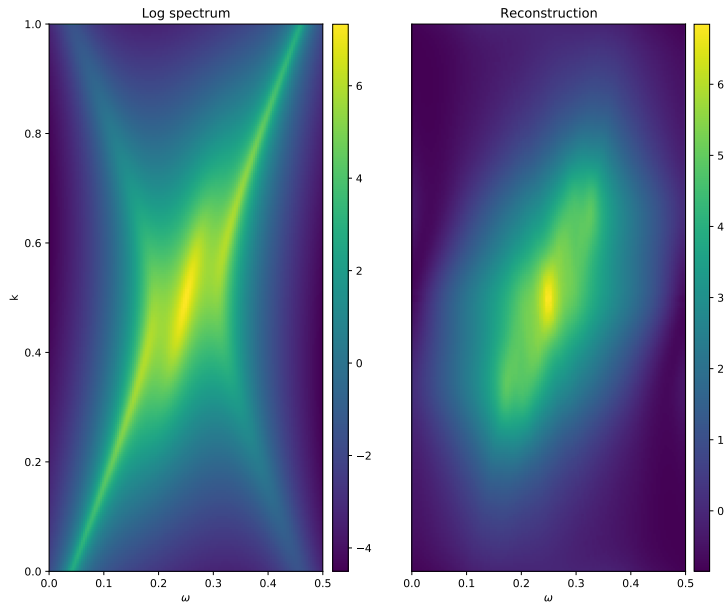
Application



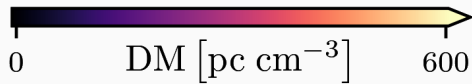
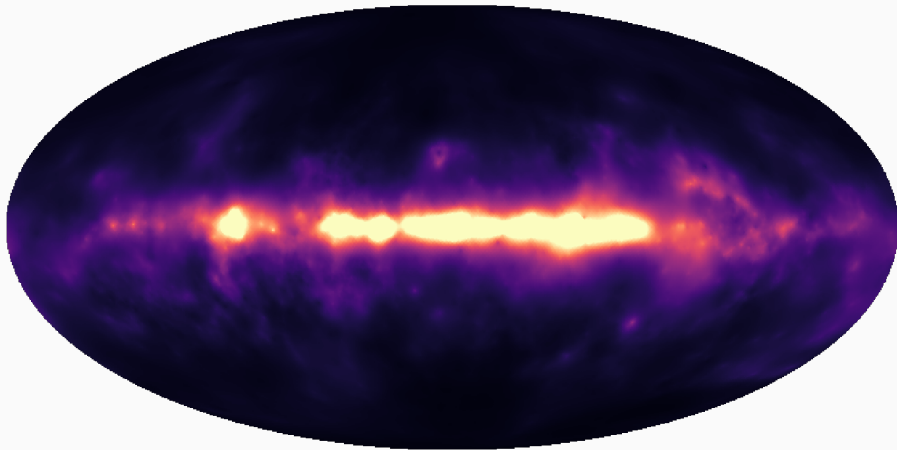
Application



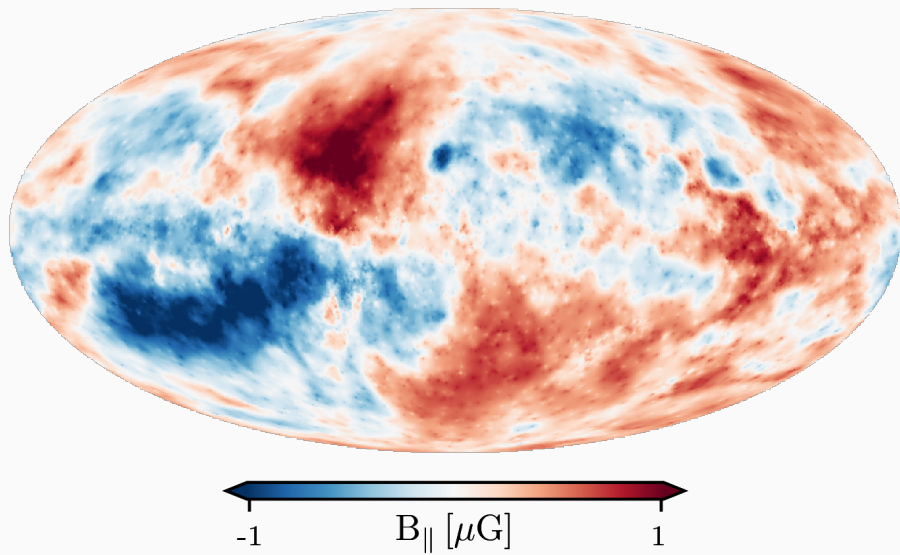
Application



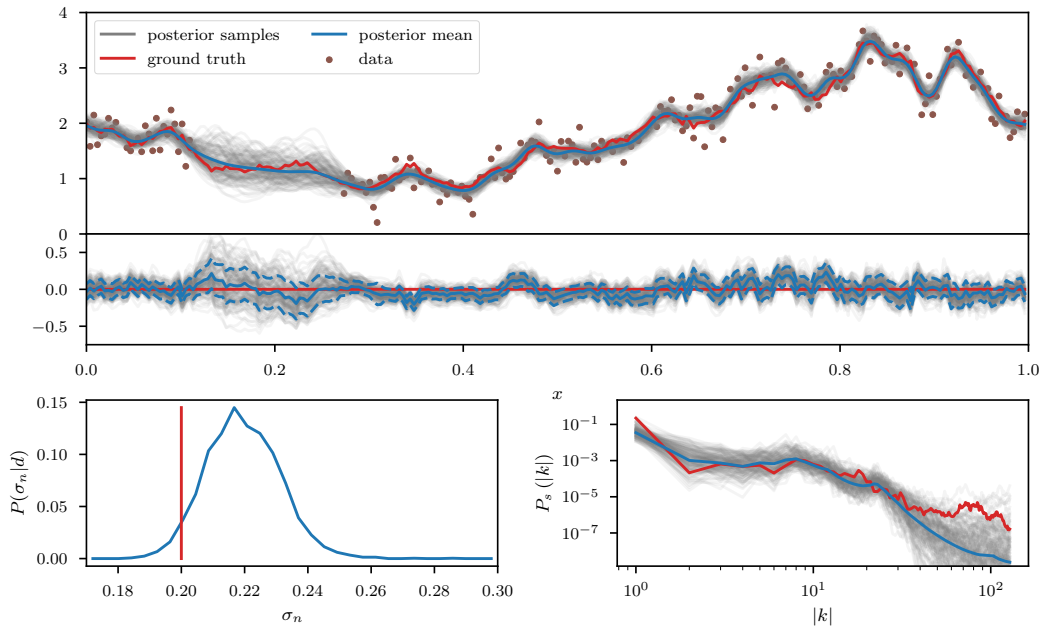
Appendix



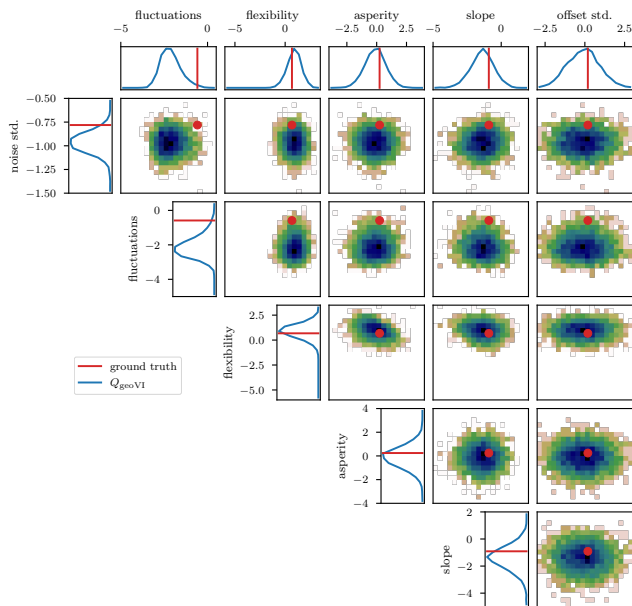
Appendix



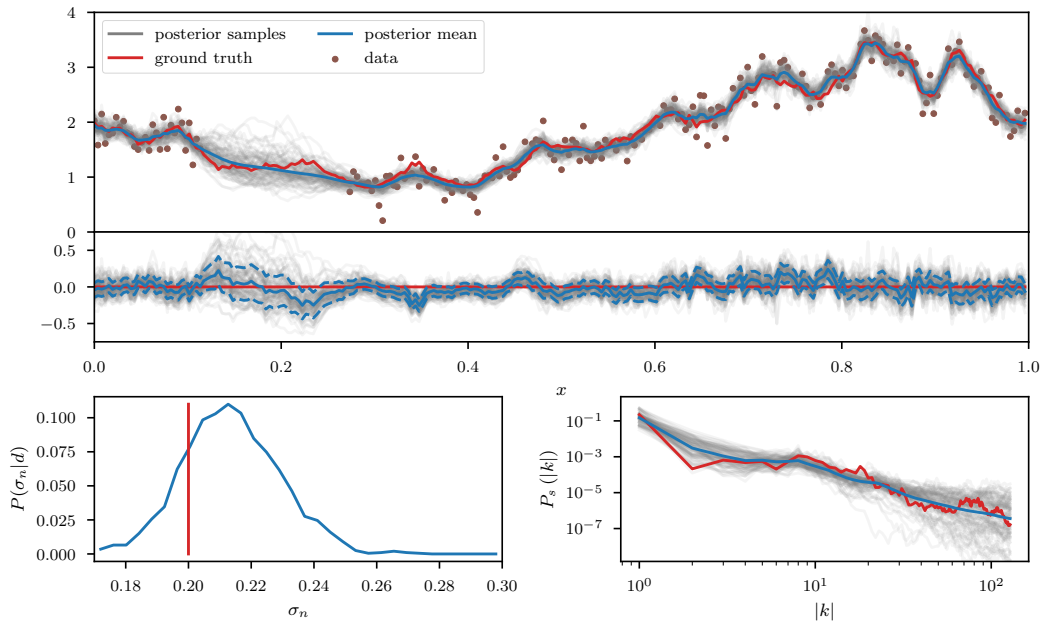
Appendix



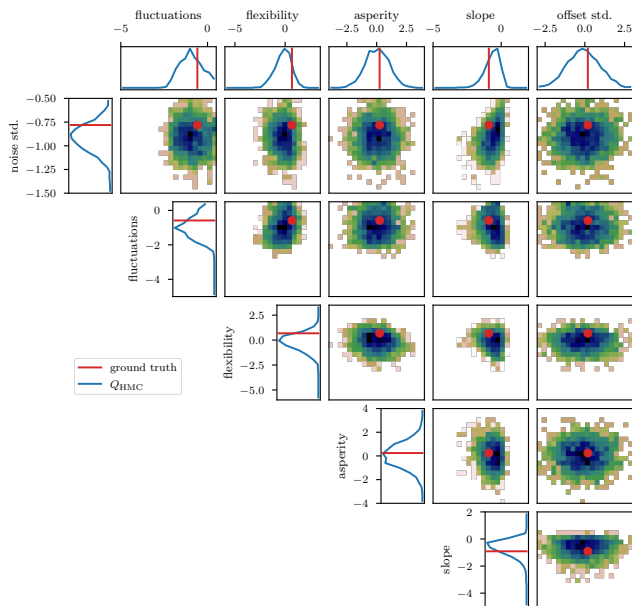
Appendix



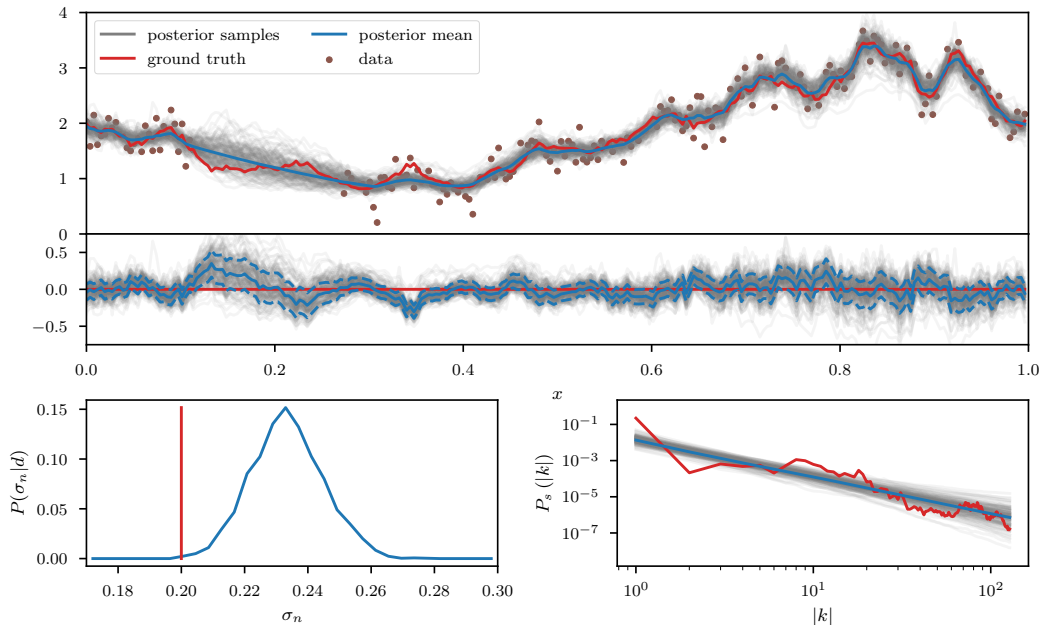
Appendix



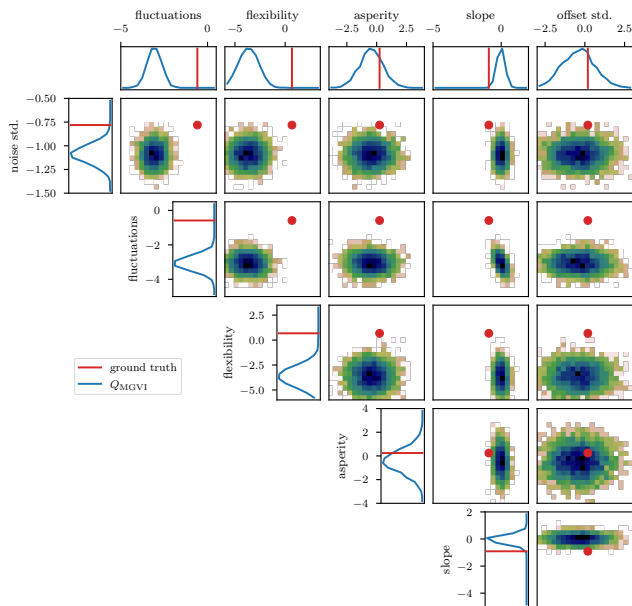
Appendix



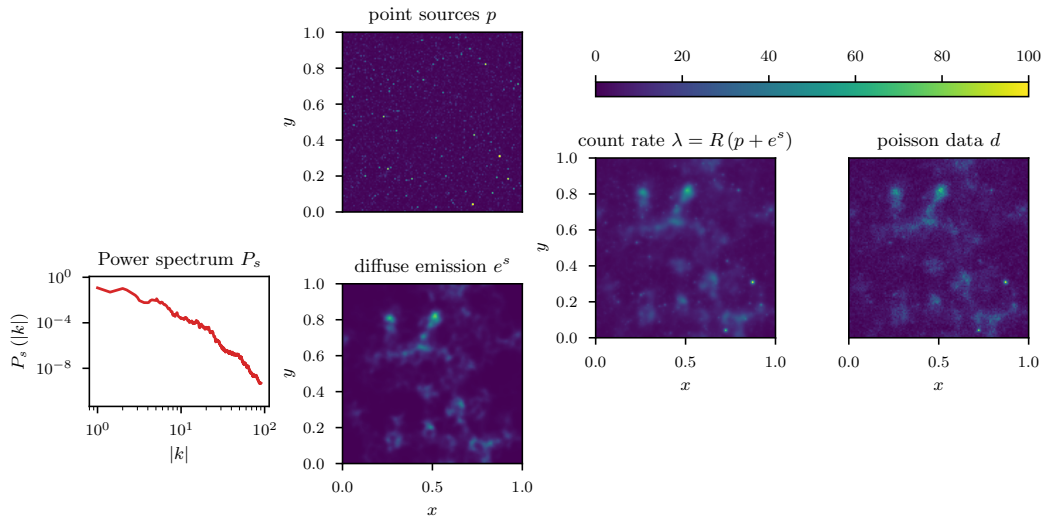
Appendix



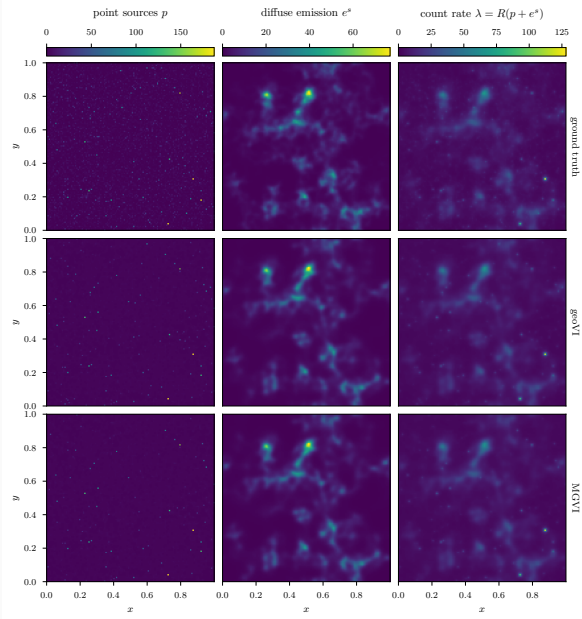
Appendix



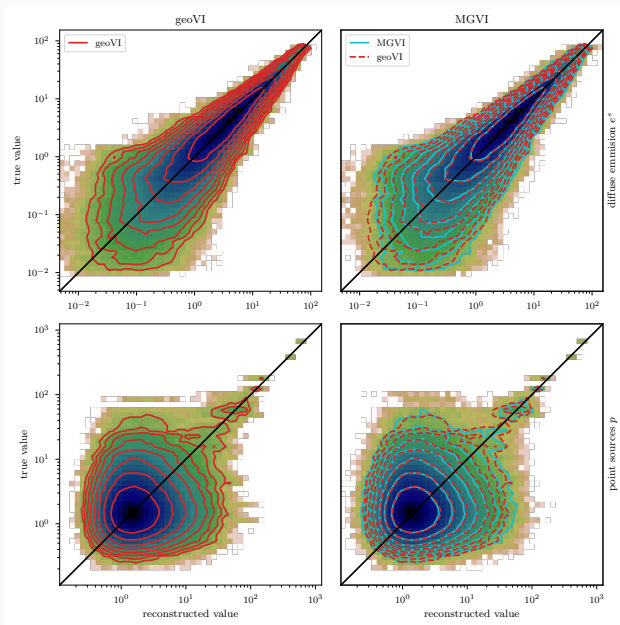
Appendix



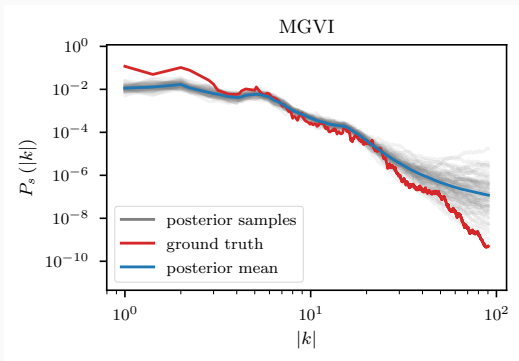
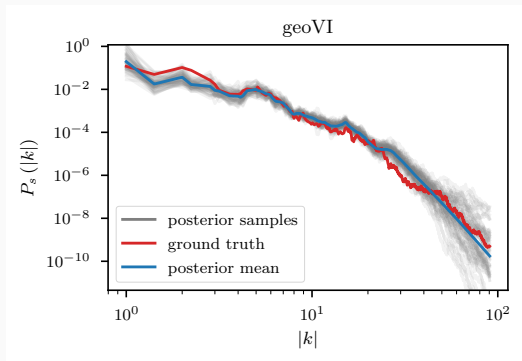
Appendix



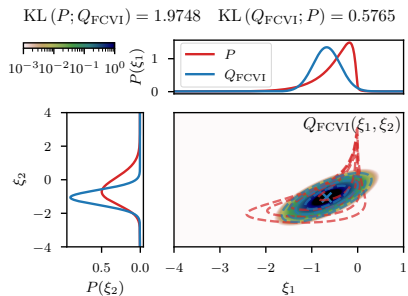
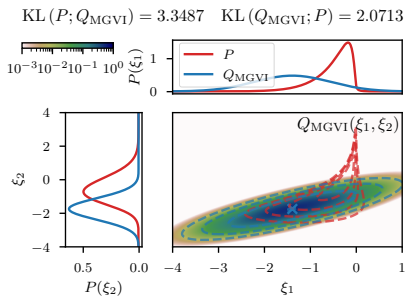
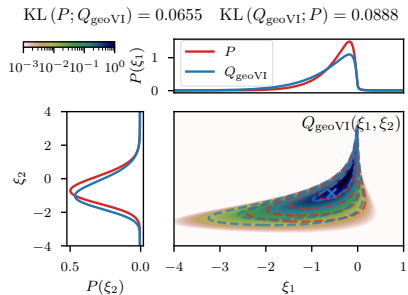
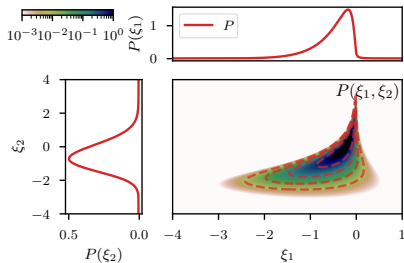
Appendix



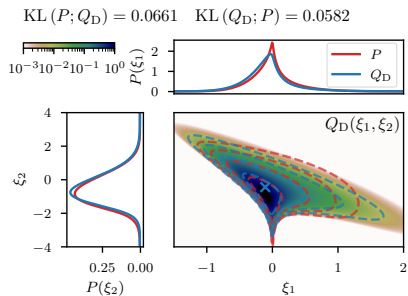
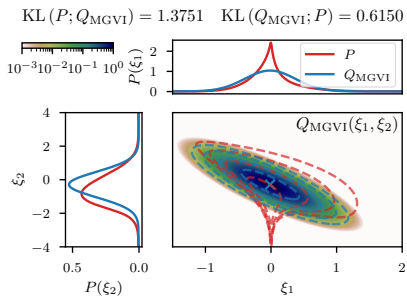
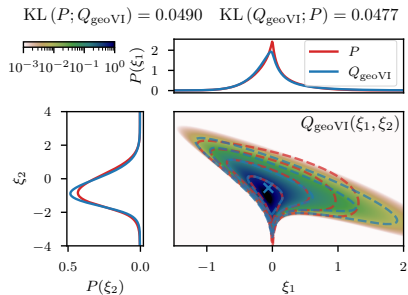
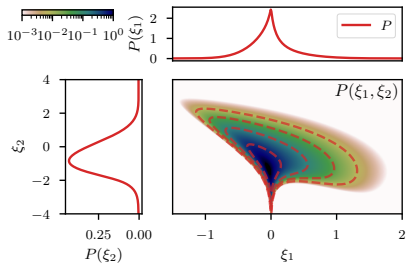
Appendix



Appendix

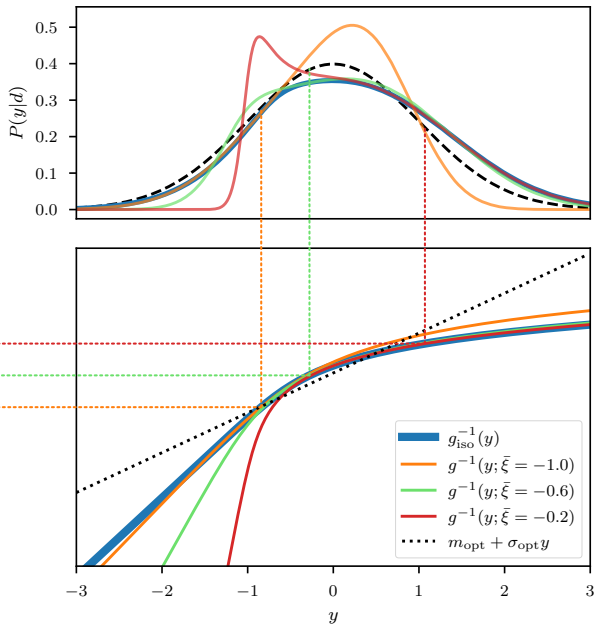
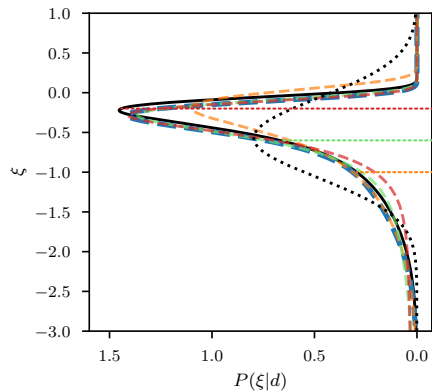


Appendix



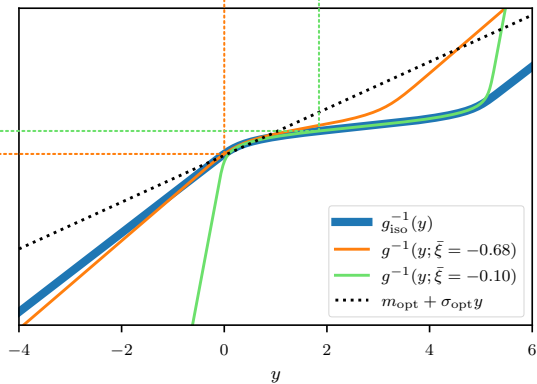
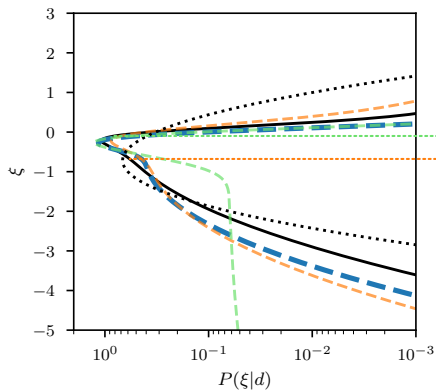
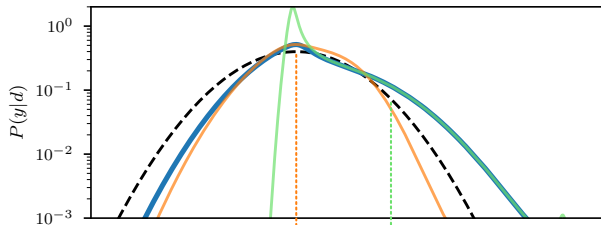
Appendix

$$\begin{aligned}
 \text{KL}(P; Q_{\text{iso}}) &= 0.0333 \\
 \text{KL}(P; Q_{\bar{\xi}=-1.0}) &= 0.0582 \\
 \text{KL}(P; Q_{\bar{\xi}=-0.6}) &= 0.0489 \\
 \text{KL}(P; Q_{\bar{\xi}=-0.2}) &= 0.1557 \\
 \text{KL}(P; Q_{\text{Normal}}) &= 0.2864
 \end{aligned}$$



Appendix

$$\begin{aligned}\text{KL}(P; Q_{\text{iso}}) &= 0.0806 \\ \text{KL}(P; Q_{\bar{\xi}=-0.68}) &= 0.0642 \\ \text{KL}(P; Q_{\bar{\xi}=-0.10}) &= 0.5656 \\ \text{KL}(P; Q_{\text{Normal}}) &= 0.1817\end{aligned}$$



Appendix

